ON RESEARCH EFFICIENCY

A MICRO-ANALYSIS OF DUTCH UNIVERSITY RESEARCH IN ECONOMICS AND BUSINESS MANAGEMENT*

L. Cherchye and P. Vanden Abeele**
Catholic University of Leuven
Campus Kortrijk and Faculty of Economics and Applied Economics

ABSTRACT

We argue that efficiency assessments of academic research should focus on *micro*-units of research production rather than on conventionally employed (aggregated) *macro*-units, and show that such a detailed analysis of research performance provides interesting insights. In addition, we propose a non-parametric methodology that is specially tailored for analyzing the productive efficiency of research: it starts from a specification of the managerial objectives of research activities while imposing minimal structure on the (typically unknown) production technology. We illustrate our points by assessing the productive efficiency of research in Economics and Business Management faculties of Dutch universities. Next to measuring productive efficiency, we look for specific patterns in efficiency distributions over universities, years and areas of specialization. In addition, we investigate the impact of external funding and of the size of research programs on academic research efficiency.

KEYWORDS: productive efficiency of research, micro-units of production, non-parametric analysis, research in economics, determinants of research efficiency

JEL CLASSIFICATION: A11, C14, C24, C61, D24

---

* We thank J. W. Meijer, L. Parsons and A. Watteyne for helpful comments and suggestions.
** Corresponding author:
address: Catholic University of Leuven, Campus Kortrijk, KTEW, E. Sabbelaan 53, B-8500 Kortrijk, Belgium;
tel: +32 (0)56 246109; **fax**: +32 (0)56 246999; **e-mail**: Laurens.Cherchye@kulak.ac.be.

1. INTRODUCTION

There is a general interest in performance assessments of research activities (often referred to as R&D), if only because they provide useful information for distributing the limited financial resources that are available, and because 'bad' performers can learn from 'good' performers. In particular the 'productive' efficiency of research has received much attention. Generally, efficient research production is not guaranteed by the usual market correction mechanisms, which calls for tools to quantify research efficiency - some even argue that an effective performance measurement system is a necessary condition for R&D productivity; see e.g. Cordero (1990). Still, the construction of research evaluation tools is no simple task, since multiple (input and output) performance dimensions should be taken into consideration.

This demand for operational evaluation tools is not exclusively concentrated in the business community. In academia, the research dimension equally gains importance, which has stimulated a number of authors to summarize academic research activity in single-valued performance measures. In particular, faculties of Economics and Business Management have been subjected to evaluation; see e.g. Kocher et al. (2001), Korhonen et al. (2002), Thursby (2000) and references therein.[1]

We contribute to that strand of literature by analyzing the productive efficiency of research in Economics and Business Management organized at Dutch universities. Our data cover the period 1996-2000 and were delivered by the universities in the context of the quinquennial assessment of university research, conducted under the auspices of the Association of Dutch Universities (VSNU). Although our specific focus is on academic research performance, we believe that our main points are also applicable to research analysis in the business sector.

Our approach deviates from the mainstream literature in at least two respects:

First, we use *research programs* as our *micro*-units of assessment, rather than comparing Economics departments or other aggregated *macro*-units.[2] Each faculty of Economics and/or Business Management harbors at least one, and generally several research programs. A 'program' is conducted by a group of researchers who join efforts and resources in order to investigate a particular theme, and in the process to educate researchers and to publish research results.

In our opinion, research programs form the natural observation units for studying academic research efficiency. In fact, concentrating on these micro-units of research production provides a more detailed insight into the productive efficiency of academic research activities. For example, micro-analysis allows us to refine our examination to specialization domains within the general field of Economics and Business Management; this can reveal efficiency differences within universities (between research units of different specialization types) and between universities (within a particular specialization area). Evidently, this information is not obtained from a (conventional) macro-analysis, as such efficiency variation will be 'obscured' when aggregating all research programs per department or faculty. This is an important point to make as the outcomes of a productive efficiency analysis can have far-reaching implications, e.g. pertaining to the allocation of financial resources, so that aggregating all research programs may generate undesirable effects.

Second, our study is innovative in terms of the methodology that we employ. Our specific methodological orientation is induced by a number of practical difficulties associated with the

---

[1] See also the recent call by the European Economic Association for studies on "Ranking Economic Departments throughout Europe" (http://www.eeassoc.org/ranking%20.html).

[2] Research *programs* should be distinguished from research *projects*; see e.g. Cooper (1996) and Griffin and Hauser (1996) for general discussions of concepts and assessment strategies.

empirical approximation of the research technology. One such difficulty is that the production of research typically involves multiple inputs and multiple outputs, which makes it problematic to use standard parametric/regression techniques. Another, more serious problem is that minimal 'engineering' knowledge is usually available about the precise interrelationship between the research inputs that are used and the research outputs that are produced.

Non-parametric efficiency analysis techniques circumvent both these problems: they allow for efficiency evaluation without necessitating the specification of a functional representation of the technology, and they naturally deal with the simultaneous occurrence of several inputs and multiple outputs.[3] These attractive features have inspired a number of authors (e.g. Kocher et al. (2001), Korhonen et al. (2002) and Thursby (2000)) to apply non-parametric techniques for assessing academic research efficiency. Still, these authors persistently start from a specification of basic properties of the production technology (e.g. pertaining to the nature of the prevalent returns-to-scale), while –to recall- minimal technological information is available in the context of research production.

In this paper, we adopt the opposite perspective: we start from an explicit characterization of the eventual objectives of the research programs while imposing the least structure on the production technology; we only use observed combinations of inputs used and outputs produced to approximate the technological possibilities. Attractively, this *dual* orientation falls in line with the seminal contributions by Afriat (1972) and Varian (1984, 1990) on non-parametric production and efficiency analysis.

Of course, objectives will vary according to the nature of each program, and the weights put on inputs and outputs is implicit and specific to each research program. Therefore, we merely specify some basic characteristics that seem acceptable for the objective of any research program. Our model then allows the managers of the research programs to define their own objective function, or even better, the method reveals the objective function which is optimal from an efficiency perspective. Putting it somewhat differently, the nonparametric methodology allows for giving each research program the *benefit-of-the-doubt* in the absence of full information about the true input and output weights.

Our 'efficiency measure' provides information on the extent to which the behavioral objectives are achieved. More specifically, it can be interpreted as an upper bound estimate for the input cost efficiency of research programs: for the given research output, it bounds the ratio of minimal cost over actual cost from above; research output is then a combination of individual outputs, which can potentially increase in one or several of its constituent components.

Next to measuring productive efficiency, we want to *explain* significant differences in research performance. In a first step, we try to discern patterns in the distribution of efficiencies over universities; we hereby also consider variation in university efficiencies over years and over specialization domains. As we try to exploit the information in the data to the fullest extent while imposing minimal non-verifiable structure on the setting under investigation, we again proceed non-parametrically and use bivariate Wilcoxon rank sum tests to address the issue. In a second step, we use multivariate Tobit regression analysis to scrutinize the particular relationship between observed efficiency on the one hand and the size of research programs and financial support by scientific research funds on the other. Our results should give us more insight into the determinants of the observed 'success' of research programs, and can provide direct policy guidelines for program managers.

[3] See e.g. Färe et al. (1994) and Cooper et al. (2000) for introductory textbooks on nonparametric production and efficiency analysis.

The remainder of this paper unfolds as follows. Section 2 introduces our efficiency assessment methodology. Section 3 discusses our selection of inputs and outputs, presents our efficiency results and compares these results with the VSNU assessment results supplied by an expert committee. Section 4 examines possible determinants of research performance. Section 5 provides some concluding discussion.

2. Measuring Productive Efficiency: Methodology

We denote the research input vector by $x \in \Re_+^l$ and the research output vector by $y \in \Re_+^m$. The set of all technologically feasible input-output combinations is the production possibility set

(1) $$T \equiv \left\{ (x, y) \in \Re_+^{l+m} \middle| x \text{ can produce } y \right\}.$$

Efficiency analysis relates research input to research output. For that purpose we need to aggregate the different components of the vectors x and y. We value total input in cost terms, i.e. we use *input cost px* (or $p \cdot x$) for given input price vector $p \in \Re_+^l$; to simplify our exposition we restrict to $px > 0$. Overall *research output* is obtained as a combination of individual outputs by using an output value function $V : \Re_+^m \to \Re_+ : y \to V(y)$, where $V(y) \geq 0 \; \forall y \in \Re_+^m$.

Efficiency Evaluation: Theoretical Approach

As indicated in the introduction, our efficiency measurement model starts from a specification of the behavioral objectives that serve as a basis for (ex post) efficiency evaluation. Loosely stated, our model assumes that the managers of research programs aim at '*providing the research output at minimal cost*'.

To translate that behavioral assumption into a formal efficiency condition, we denote the minimal cost associated with a particular research output (dependent on $y \in \Re_+^m$) under price vector $p \in \Re_+^l$ and production possibility set T by

(2) $$C^{T,V}(y; p) \equiv \min_{(x', y') \in T} \left\{ px' \middle| V(y') \geq V(y) \right\}.$$

Our formal efficiency condition can be stated as follows: for given price vector $p \in \Re_+^l$, select $(x, y) \in T$ such that

(3) $$px = C^{T,V}(y; p).$$

Using that $px > 0$, this efficiency condition can be checked by means of the efficiency measure

(4) $$\varphi^{T,V}(x, y; p) \equiv \frac{C^{T,V}(y; p)}{px}; \; 1 \geq \varphi^{T,V}(x, y; p) \geq 0.$$

This measure tells us how 'close' observed production behavior is to efficient behavior, where 'closeness' is measured in terms of input cost. More specifically, the interpretation of $\varphi^{T,V}(x, y; p)$ is as follows:

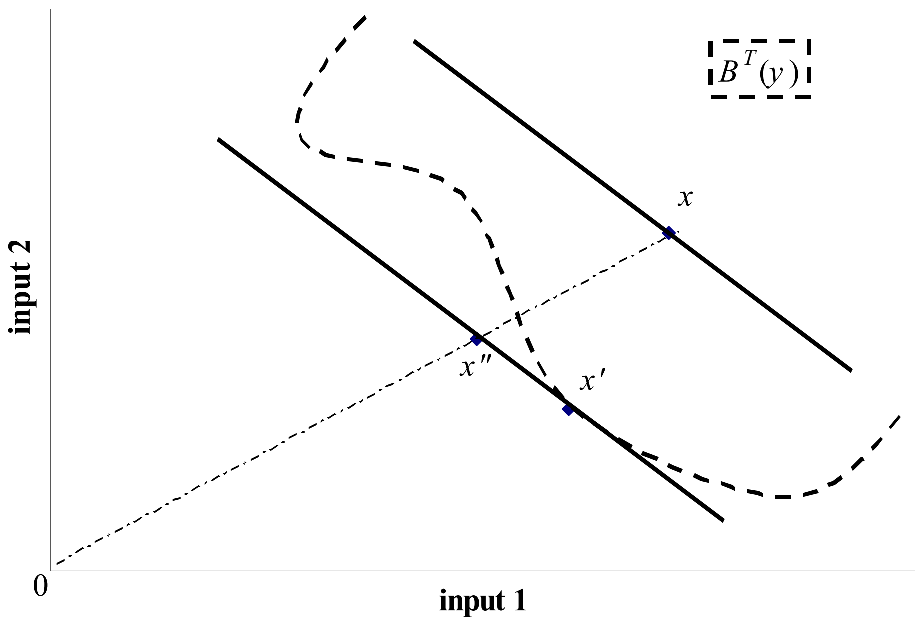
a. if $\varphi^{T,V}(x,y;p)=1$, then the (necessary and sufficient) efficiency condition (3) is met;

b. if $\varphi^{T,V}(x,y;p)<1$, then condition (3) is not met; in that case, the measure $\varphi^{T,V}(x,y;p)$ captures the ratio of minimal cost over actual cost, or: it measures the degree to which actual cost should be reduced to be cost efficient for the given research output.

The focus on input performance seems well justified: typically, research managers only control the cost-generating inputs while the research outputs are at least partly determined exogenously, and it seems intuitive to measure research performance only in terms of controllable dimensions.

To conclude our discussion, we graphically illustrate $\varphi^{T,V}(x,y;p)$ in Figure 1, which presents the case of 2 inputs. In that figure $B^T(y)=\{x'|(x',y')\in T \wedge V(y')\geq V(y)\}$ contains all input vectors that produce at least the research output $V(y)$. Let the price vector p for (x,y) correspond to the slope of the bold iso-cost line through x. The measure $\varphi^{T,V}(x,y;p)$ then compares the cost level associated with x with the minimum cost level over the set $B^T(y)$, which is achieved in x'. More specifically, $\varphi^{T,V}(x,y;p)$ is computed as the ratio of minimum cost over actual cost, which equals 0x"/0x.

Figure 1: efficiency measurement – theoretical case



EFFICIENCY EVALUATION: EMPIRICAL APPROACH

If complete information about the production possibility set (T), the input prices (p) and the output value function (V) were available, we could readily test the *necessary and sufficient* (*theoretical*) condition (3) and compute the efficiency measure $\varphi^{T,V}(x,y;p)$. However, in many cases, as in our application below, such complete information is not available. Therefore, we proceed by constructing a *necessary* (*empirical*) condition for efficient behavior by gradually weakening the informational requirement.

a. *Production possibility set T is not observed*

To deal with incomplete technological information, the non-parametric orientation suggests to start from the set of n observed input-output vectors $S \subseteq T$ ($card(S) = n$); see e.g. Varian (1984). Using S instead of T in (4) we get the efficiency measure

$$(5) \qquad \varphi^{S,V}(x, y; p) = \frac{C^{S,V}(y; p)}{px}.$$

Given that $S \subseteq T$, we have $1 \geq \varphi^{S,V}(x, y; p) \geq \varphi^{T,V}(x, y; p)$, or: $\varphi^{S,V}(x, y; p) = 1$ is a necessary condition for $\varphi^{T,V}(x, y; p) = 1$ (i.e. for (3)).

b. *Price vector p is not observed*

In line with our search for 'necessary' efficiency conditions, we use 'most favorable' input prices (or 'shadow' prices) to assess the efficiency of input-output combinations, i.e. we apply 'benefit-of-the-doubt' pricing in the absence of full price information. This yields the efficiency measure

$$(6) \qquad \rho^{S,V}(x, y) \equiv \max_{p \in \Re^l_+ : px > 0} \varphi^{S,V}(x, y; p) = \max_{p \in \Re^l_+ : px > 0} \frac{C^{S,V}(y; p)}{px}.$$

It is easy to verify that $1 \geq \rho^{S,V}(x, y) \geq \varphi^{S,V}(x, y; p) \quad \forall p \in \Re^l_+ : px > 0$, or: $\rho^{S,V}(x, y) = 1$ is a necessary condition for $\varphi^{T,V}(x, y; p) = 1 \quad \forall p \in \Re^l_+ : px > 0$ (and thus for (3)).

Using (2), we can also write $\rho^{S,V}(x, y)$ as

$$(7) \qquad \rho^{S,V}(x, y) = \max_{p \in \Re^l_+ : px > 0} \left[\min_{(x', y') \in S} \left\{ \frac{px'}{px} \middle| V(y') \geq V(y) \right\} \right].$$

Hence, we still need information on V. This is discussed next.

c. *Value function V is not observed*

We 'reconstruct' the non-observed output value function axiomatically, i.e. we start from minimal assumptions about the function V that seem generally acceptable. In particular, we impose that V is monotonically increasing in outputs, i.e. $y' \geq y \Rightarrow V(y') \geq V(y)$.

This assumption makes for $(x, y) \in S : \{(x', y') \in S \mid y' \geq y\} \supseteq \{(x', y') \in S \mid V(y') \geq V(y)\}$. Hence, we obtain the efficiency measure

$$(8) \qquad \theta^{S}(x, y) \equiv \max_{p \in \Re^l_+ : px > 0} \left[\min_{(x', y') \in S} \left\{ \frac{px'}{px} \middle| y' \geq y \right\} \right],$$

or, using $px > 0$ and u to denote the minimal cost level,

(9) $$\theta^S(x,y) = \max_{p \in \Re^l_+} \{ u \mid px = 1; u \le px' \quad \forall (x',y') \in S : y' \ge y \}.$$

We easily find that $1 \ge \theta^S(x,y) \ge \rho^{S,V}(x,y)$, or: $\theta^S(x,y) = 1$ is a necessary condition for $\rho^{S,V}(x,y) = 1$ (and thus for (3)).

We end up with $\theta^S(x,y)$ as an efficiency measure that uses minimal (available) information. This measure has a direct interpretation as an upper bound for $\varphi^{T,V}(x,y;p)$, i.e. $\theta^S(x,y)$ captures the minimally feasible cost reduction for the given research output.

To conclude our methodological discussion, we point out two interesting features of the measure $\theta^S(x,y)$. First, the values of $\theta^S(x,y)$ can be computed by simple linear programming after identifying the set $D^S(y) \equiv \{(x',y') \in S \mid y' \ge y\}$ in a first step. The fact that merely linear programming is needed for the computation of the efficiency measure (after a trivial check of output dominance) is generally attractive for practical applications.

A second note applies to the dual formulation of (9):

(10) $$\theta^S(x,y) = \min_{\lambda(x',y') \ge 0 \quad \forall (x',y') \in D^S(y)} \left\{ \theta \,\middle|\, \theta x \ge \sum_{(x',y') \in D} \lambda_{(x',y')} x'; \sum_{(x',y') \in D} \lambda_{(x',y')} = 1 \right\}.$$

This expresses $\theta^S(x,y)$ as the Debreu-Farrell efficiency gauge (see Debreu (1951); Farrell (1957)) calculated with respect to the empirical production possibility set

(11) $$T^E(S) = \left\{ (x,y) \in \Re^{m+n}_+ \,\middle|\, \begin{array}{l} x \ge \sum_{(x',y') \in S} \lambda_{(x',y')} x'; \sum_{(x',y') \in S} \lambda_{(x',y')} = 1; \\ \lambda_{(x',y')} y' \ge \lambda_{(x',y')} y \wedge \lambda_{(x',y')} \ge 0 \quad \forall (x',y') \in S \end{array} \right\},$$

which was first introduced by Hanoch and Rothschild (1972).

The Debreu-Farrell measure in (10) captures the maximum equiproportionate input reduction for given output within $T^E(S)$. This efficiency evaluation model was already proposed by Bogetoft (1996), but from a very different motivation. More specifically, Bogetoft motivates this efficiency assessment model from properties of the production possibility set, while our reasoning starts from a specific characterization of the objectives of research programs and does not use any technological assumption apart from $S \subseteq T$. The above reasoning thus provides an economic (as opposed to 'engineering') interpretation of the model suggested by Bogetoft.

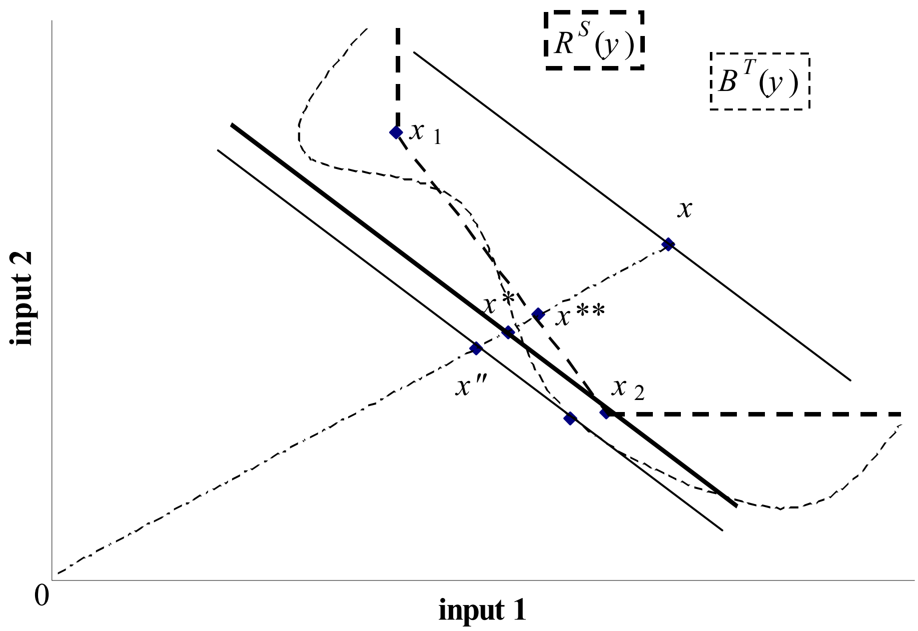
We graphically illustrate our discussion in Figure 2, which continues our previous example. In this case we do no longer observe the set $B^T(y)$. We approximate that set by the set of observed input vectors that produce at least the output y, i.e. $\{x' \mid (x',y') \in S \wedge y' \ge y\}$, which we assume to be $\{x, x_1, x_2\}$ in our example. Obviously, $\{x, x_1, x_2\} \subseteq B^T(y)$.

If we knew the true prices associated with the vector x, we would measure productive efficiency as the ratio 0x*/0x, using that x_2 is cost minimizing over $\{x, x_1, x_2\}$; notice that 0x*/0x > 0x''/0x. If the true prices are unknown, we use most favorable prices to evaluate the efficiency of (x,y). These most favorable prices correspond to the slope of the line segment $x_1 x_2$, i.e. the relative prices under which both x_1 and x_2 are cost minimizing over $\{x, x_1, x_2\}$. The resultant efficiency estimate for x equals 0x**/0x. Clearly, 0x**/0x > 0x*/0x, which demonstrates the 'benefit-of-the doubt' pricing that underlies $\theta^S(x, y)$.

This example also demonstrates the above dual interpretation of the efficiency measure: the set $R^S(y) = \{x' | (x', y') \in T^E(S) \wedge y' \geq y\}$, and the ratio 0$x$**/0$x$ is (dually) obtained as the Debreu-Farrell efficiency gauge computed with respect to that set. More specifically, 0x**/0x captures the maximal equiproportionate input reduction within $R^S(y)$.

As a final note, observe that the more information we have (i.e. the more input-output combinations we observe), the better the empirical approximation of the set $B^T(y)$ will be, and so the better the upper bound efficiency estimates will approximate the true efficiency measure $\theta^S(x, y)$. This immediately highlights the usefulness of large data sets for obtaining high-quality efficiency estimates.

Figure 2: efficiency measurement – empirical case



INTERPRETING 'INEFFICIENCY'

It is important to emphasize that the computed efficiency values can have different interpretations. Their usual interpretation is that revealed inefficiency indicates truly inefficient behavior, e.g. due to agency problems caused by imperfect monitoring of the research production process. However, revealed inefficiency may also indicate ill-specification of the efficiency evaluation model. Since our approach imposes minimal structure on the production technology, such ill-specification should pertain to our assumptions about the research objectives. In this respect, the cost minimization assumption seems reasonably acceptable. However, on the output side we do not account for uncertainty in the production of research outcomes, an arguable assumption in the setting under investigation. It is, hence, to be kept in mind that our efficiency results give an indication of *ex post* efficiency.

Still, we believe that our focus on ex post efficiency is well-grounded. After all, academic research is conventionally evaluated in terms of ex post results, also in subjective assessments such as that conducted under the auspices of the VSNU. In addition, differences in ex post efficiency at least give us an indication about how different research programs deal with the ex ante uncertainty, which in turn provides information on the quality of these research programs. This signaling interpretation certainly applies when differences in research performance occur systematically.

Two alternative interpretations of the reported inefficiency values pertain to the data collection process. First, relevant input or output dimensions may be omitted. Second, our model does not account for errors-in-the-data. Obviously, omitted performance dimensions and measurement errors may distort the individual (ex post) performance results, so that conclusions drawn from these individual efficiency values could be misleading. More specifically, these phenomena can affect the observed efficiency differences and even the ranking of research programs.

These considerations suggest that, on the level of individual research programs, the efficiency measures should primarily be used as screening devices and for attention-direction. In-depth investigation of seemingly problematic research programs is necessary before drawing conclusions regarding research performance.

Still, we can reasonably expect that the impact of omitted performance dimensions and measurement errors will work in a parallel way when comparing efficiency results for specific categories of research programs. Therefore, in our remaining discussion, we will primarily concentrate on systematic efficiency differences between groups of research programs rather than on efficiency differences between individual programs. In addition, we will employ parametric regression techniques, which by construction account for outlier behavior and so mitigate the potential impact of omitted input or output dimensions and errors-in-the-data.

3. MEASURING PRODUCTIVE EFFICIENCY: EMPIRICAL RESULTS

Our input and output data are taken from the 'Quality Assessment Reports on Research 1996-2000', delivered by each Dutch university in the context of the quinquennial VSNU assessment. These extensive self-assessments contain both quantitative and qualitative information on inputs, throughputs and outputs of academic research. The data in these VSNU reports are detailed, compared to figures generally available in this domain; also, the data are relatively well standardized and have been subjected to some scrutiny for correctness and consistency, which at least partly obviates the above interpretation of reported inefficiency as pertaining to the quality of the data collection process.

We have data for 79 research programs organized at 8 universities: Erasmus University Rotterdam (EUR), Universiteit van Amsterdam (UvA), Vrije Universiteit Amsterdam (VU), Universiteit Maastricht (UM), Tilburg University (KUB), Wageningen University (WUR), University of Groningen (RUG) and University of Nijmegen (KUN).[4] These data are available for 5 years (1996 to 2000). We thus have data for 495 observations (i.e. 79 research programs over 5 years) in total, which makes a data set that is significantly larger than usual considered in this kind of applications.

A preliminary note applies to our efficiency estimates. Abstracting from omitted performance dimensions and measurement errors (discussed above), our estimates merely provide (upper bound) proxies for the true efficiency values. This may well affect our results, as the computed efficiency differences and rankings may differ from the true efficiency differences and rankings. This limitation should be kept in mind when interpreting the results reported below. Still, we strongly believe that our choice for this second-best route is well defensible, especially since the first-best route is simply not feasible due to limited information about the technology, the input prices and the output value functions. Also recall that the relatively large size of our data set benefits the quality of our efficiency estimates.

Before presenting our efficiency results, we discuss the input and output data originally reported in the assessment reports, and motivate the input and output selection used in our efficiency analysis.

INPUT DATA

The assessment reports contain data for 6 inputs (expressed in full time equivalents): PhD candidates WP1 (I1); other research input (postdoctoral fellows, professors, associate professors and other senior staff) WP1 (I2); PhD candidates WP2 (I3); other research input WP2 (I4); PhD candidates WP2 (I5); other research input WP3 (I6). The distinction between the categories WP1, WP2 and WP3 pertains to the source of funding; WP1 means that funding is from internal sources, while WP2 and WP3 refer to alternative external sources, respectively scientific research funds and contract research grants.

It is to be noted that different universities use different allocation methods for determining the actual contribution of researchers to the programs. It could be argued that this introduces heterogeneity over inputs. Still, we use the input data originally provided by the universities. A first reason is that it is not always possible to 'homogenize' input values from the information that is available. More importantly, universities can allocate time differently over research and e.g. teaching, and hence researchers can indeed contribute differently to the research program with which they are associated; in addition, WP1, WP2 and WP3 researchers can allocate a different proportion of their working time to research. We believe that universities are best placed to choose the allocation method that is most appropriate. Finally, it is worth to mention that the variation in allocation methods is not very substantial.

OUTPUT DATA

The reports provide data for 8 outputs (expressed in total numbers): doctoral dissertations (O1); refereed articles in international journals (O2); non-refereed articles in international journals (O3); ((co-)edited) books (O4); chapters in books and proceedings (O5); refereed articles in Dutch journals (O6); non-refereed articles in Dutch journals (O7); professional publications and scientific reports for third parties (O8).

[4] The reports actually contain data for 2 additional research programs. One has been excluded from our evaluation because it was not assessed by the VSNU, and the other has been excluded because data were only available for the years 1999 and 2000.

We do not use the data for O3, O6, O7 and O8 in our empirical exercises below. In contrast to the other outputs, these outputs are rather vaguely defined and, apparently, the categories are quite differently interpreted by different research programs. Finally, these outputs seem to be of minor importance for assessing the performance of research programs.

As for O2, we exclude (non-refereed) 'book reviews' from the list of 'refereed articles in international journals'. Next, in view of the importance that is generally attributed to this output, we make a further distinction between publications in international 'top' journals (O2a) and publications in 'other' international journals (O2b = O2-O2a); see appendix A for the list of international top journals that we use.[5] In all our exercises we use both O2 and O2a as outputs. By double counting the output O2a, we impose that for one program to achieve a higher value level than another program in terms of the output 'international journal articles' a *sufficient* condition is that the first program is associated with a higher value for O2 (= O2a+O2b) *and* a higher value for O2a, or: for the former program to be a possible comparison partner for the latter program under the limited information that is available, it should produce more articles in refereed journals *and* more articles in international 'top' journals. This procedure is consistent with our focus on *necessary* efficiency conditions

Finally, as we concentrate on 'internationally oriented' academic research, we restrict attention to publications in an 'internationally accessible language' (i.e. English, German or French) for O4 and O5.

EFFICIENCY RESULTS

We compute efficiency values for three different selections of inputs and outputs. In the main study (yielding efficiency results 'eff1'), we use data for two inputs: 'PhD candidates' and 'other research input', where we sum the values for WP1, WP2 and WP3 (i.e. our inputs are I1+I3+I5 and I2+I4+I6, respectively). We relate these two inputs to three outputs: 'total number of doctoral dissertations' (O1), 'total number of refereed articles in international top journals' (O2a) and 'total number of refereed articles in international journals in general' (O2=O2a+O2b).

To explicitly account for the fact that inputs in one particular year may generate outputs in subsequent years, we relate the output in a specific year to the sum of the inputs used in that same year and the inputs used in the two preceding years. This makes that we end up with 237 observations, i.e. 79 research programs over 3 years (1998-2000).

A different selection of inputs and/or outputs may be more appropriate, but data availability and the desirability of homogenous inputs and outputs limit our options. Still, to investigate the sensitivity of our results with respect to the selected input and output dimensions, we also compute efficiency estimates (labeled 'eff2' and 'eff3') for two alternative selections of inputs and outputs, and compare these with our original results. First, on the input side we account for the fact that WP1, WP2 and WP3 inputs may be unequally weighted: we distinguish between inputs according to their funding (WP1, WP2 and WP3), to end up with six inputs (I1, I2, I3, I4, I5 and I6), while holding on the original set of three outputs. Next, on the output side we consider the possibility that other outputs than those used

[5] For our selection of 'top journals', we start from the JEL-based journal weights (ranging from 1 to 5) employed by L. Bauwens, B. De Borger, P. De Grauwe and F. Heylen to rank Belgian economists and economic research institutes (see http://www.core.ucl.ac.be/econometrics/Bauwens/rankings/rankings.htm); except for some marginal modifications, we label as top journal every journal that gets a weight of 4 or 5 in that list. Evidently, our specific selection of top journals will affect at least to some extent our efficiency results. Still, we believe that our results are fairly robust with respect to other frequently used top journal lists, as these lists can be expected to overlap largely with our list.

for the construction of eff1 can play a role: we add a fourth output to the original selection, viz. books and chapters in books that are written in an internationally accessible language (O4+O5), while we stick to the original set of two inputs. In both these exercises, we correct for intertemporal effects in the same way as in our original analysis, so that we twice have 237 observations.

A survey of our efficiency results is presented in Table 1, which gives average efficiency values ('eff1', 'eff2' and 'eff3') for the 8 universities and the years 1998, 1999 and 2000.[6] We find that the results for eff1, eff2 and eff3 are largely parallel. For example, when comparing the different universities, we find that Tiburg University (KUB) and Wageningen University (WUR) are on top for all three efficiency measures. Vrije Universiteit Amsterdam (VU) follows on the third place for eff1 and eff3, and Erasmus University Rotterdam (EUR) is ranked third for eff2. University of Groningen (RUG) is situated somewhere in the middle of the ranking for eff1, eff2 and eff3. Universiteit van Amsterdam (UvA) and Universiteit Maastricht (UM) perform generally poorly in terms of productive efficiency when compared to the other universities; they are ranked last but one for at least one of our three measures. Finally, all three measures agree on University of Nijmegen (KUN) as the least efficient university on average.

Next, we consider efficiency trends over time. In view of the incentives provided by the previous VSNU assessment, which was held in 1995, we can expect efficiency to be gradually increasing over the consecutive years. We indeed find that average efficiency is substantially higher in 1999 than in 1998 for eff1, eff2 and eff3. However, this is followed by a rather drastic efficiency decrease in 2000. We cannot readily rationalize this surprising finding. At least, it suggests that it is generally useful to confront opinion with measurement (and *vice versa*).

The results in Table 1 provide some first insights into university efficiencies and distribution of efficiency values over years. Still, the fact that the reported standard deviations (see "st.dev.") are of considerable magnitude indicates that a mere comparison of average efficiencies may be misleading. In the next section, we employ more robust tests for addressing these types of questions.

To conclude this section, we compare our efficiency values with the 'scores' reported in the VSNU assessment that is based on the same input and output data. This assessment by an expert committee provides scores from 1 (poor) to 5 (excellent) for 4 criteria: quality, productivity, relevance and long-term viability. At least we would expect that our productive efficiency values are strongly and positively correlated with the VSNU scores for productivity.[7]

The relevant correlation coefficients are reported in Table 2.[8] Comparison of the computed efficiency values with the VSNU scores confirms that eff1, eff2 and eff3 have the highest correlation with the VSNU-scores for 'productivity', where the coefficient associated with eff1 is higher than that corresponding to eff2 and eff3. As for the other VSNU indicators, our three efficiency measures correlate relatively strongly with the VSNU values for 'quality' and 'viability', while correlation with the VSNU scores for 'relevance' is fairly low. This last observation should not be very surprising given that relevance (defined as use for professional

[6] Efficiencies of individual research programs are reported in appendix B.
[7] Since the VSNU scores provided by the expert committee for research programs in Business Management were not available, the results in Table 2 only pertain to research programs in Economics.
[8] We only report Pearson correlation coefficients for the sake of brevity; Spearman rank correlation results are very similar to the Pearson results.

and policy purposes) has far less to do with productive efficiency than do quality and viability.

Finally, the results in Table 2 reveal that, even though the correlation with the VSNU productivity scores is indeed outspokenly positive, it is far from perfect. Hence, there might be important deviations for individual research programs between our efficiency results and the outcomes of the VSNU assessment. This can be interpreted as revealing a need to relate output to input in a formal model when evaluating the 'operational efficiency' of research programs. As a main conclusion, we believe that the results in Table 2 at least suggest that it is generally interesting to complement expert assessment with 'objective' measurement results (and vice versa).

Table 1: average research program efficiency and standard deviations; universities and years

		eff1					eff2					eff3				
		1998	**1999**	**2000**	**total**	**ranking**	**1998**	**1999**	**2000**	**total**	**ranking**	**1998**	**1999**	**2000**	**total**	**ranking**
EUR	average	62,28%	73,78%	56,38%	64,15%	4	77,60%	85,97%	75,12%	79,56%	3	68,56%	82,91%	64,58%	72,02%	4
	st.dev.	35,97%	30,66%	32,52%	33,36%		33,91%	21,70%	32,24%	29,64%		35,99%	28,22%	29,59%	31,92%	
UvA	average	40,96%	60,65%	53,23%	51,61%	6	65,53%	75,62%	77,58%	72,91%	5	54,44%	72,60%	59,38%	62,14%	7
	st.dev.	30,33%	30,94%	30,12%	30,83%		30,57%	30,38%	29,28%	29,82%		34,68%	31,59%	31,23%	32,67%	
VU	average	65,04%	63,98%	63,76%	64,26%	3	71,30%	67,67%	67,83%	68,93%	6	72,71%	75,73%	81,82%	76,75%	3
	st.dev.	34,08%	29,04%	37,61%	32,79%		34,14%	29,19%	38,92%	33,36%		32,40%	27,39%	29,67%	29,28%	
UM	average	51,10%	57,13%	46,25%	51,49%	7	61,17%	76,84%	63,96%	67,32%	7	64,27%	74,14%	52,25%	63,55%	5
	st.dev.	32,83%	26,86%	28,80%	28,80%		29,72%	23,86%	31,36%	28,25%		25,55%	26,61%	24,08%	26,08%	
KUB	average	81,14%	89,21%	66,03%	78,80%	1	89,63%	97,31%	82,14%	89,69%	1	89,24%	90,31%	73,70%	84,42%	2
	st.dev.	31,46%	14,26%	30,16%	27,26%		29,57%	8,08%	26,06%	23,19%		17,79%	12,92%	28,85%	21,56%	
WUR	average	67,87%	87,67%	70,00%	75,18%	2	88,05%	89,47%	79,42%	85,65%	2	77,08%	99,81%	85,35%	87,41%	1
	st.dev.	26,30%	27,46%	33,93%	29,36%		18,21%	27,85%	35,23%	26,93%		26,02%	0,50%	26,17%	22,39%	
RUG	average	66,83%	60,88%	47,41%	58,37%	5	81,59%	82,83%	63,87%	76,10%	4	72,84%	63,10%	51,29%	62,41%	6
	st.dev.	28,40%	28,48%	31,60%	28,97%		29,00%	29,50%	40,13%	32,50%		27,69%	31,28%	31,14%	29,67%	
KUN	average	42,55%	54,67%	37,41%	44,88%	8	57,35%	70,51%	59,61%	62,49%	8	57,81%	59,85%	67,51%	61,72%	8
	st.dev.	40,54%	29,37%	12,64%	24,41%		60,31%	41,70%	3,13%	33,42%		27,54%	22,05%	1,58%	16,44%	
total	average	60,14%	69,59%	56,93%	62,22%		74,72%	81,29%	73,11%	76,37%		69,36%	79,25%	66,82%	71,81%	
	st.dev.	33,44%	29,16%	31,72%	31,81%		31,62%	25,93%	31,97%	30,05%		31,22%	26,89%	29,74%	29,70%	
	rank	2	1	3			2	1	3			2	1	3		

Table 2: comparison with VSNU assessment results

	correlation matrix						
	eff1	**eff2**	**eff3**	**quality**	**productivity**	**relevance**	**viability**
eff1	100,00%						
eff2	75,99%	100,00%					
eff3	84,70%	62,55%	100,00%				
quality	37,78%	30,18%	26,96%	100,00%			
productivity	53,61%	41,91%	45,99%	62,01%	100,00%		
relevance	-1,55%	3,62%	6,61%	14,55%	17,32%	100,00%	
viability	32,17%	35,40%	31,95%	64,50%	62,46%	35,52%	100,00%

4. EXPLAINING PRODUCTIVE EFFICIENCY

We explain productive efficiency by relating perceived efficiency differences to alternative factors that can be conceived as determining research performance. In the previous section, we conducted some preliminary exercises that attempted to single out the impact of the university environment, and that aimed at distinguishing efficiency variation over the consecutive years. However, these tentative results had to be interpreted with care given the high variance of the efficiency values.

We intensify our search for determinants of research performance in the current section, now employing analytical tools that explicitly account for the variance of the efficiency distribution. In a first step we analyze whether there are significant differences over universities. Again, we look for patterns over time, but now we also try to recognize differences over specialization types. Next, we examine the impact of 'controllable' dimensions such as the size of research programs and the degree of external funding of scientific research (see the WP2 inputs). We only report results based on the eff1 values for the sake of brevity; the results associated with the eff2 and eff3 values are broadly similar.

We note at the outset that the analytical tools employed below require *stricto sensu* that the efficiency estimates be independently distributed. This assumption may be criticized as the input values for the years 1998, 1999 and 2000, which are used for computing the efficiency values, are interdependent by construction for each research program (see Section 3). In addition, and probably more importantly, efficiency values are obtained from comparison with a production possibility set that is constructed from a common set of reference units, i.e. the observed set of research programs. Still, consistency results that have been established for non-parametric efficiency analysis models similar to the one applied here suggest that this interdependency problem diminishes for large samples (see e.g. Banker (1993), and Simar and Wilson (2000)). In this respect, it is worth to point out that our sample size (237 observations) is significantly above the usual size in empirical applications of non-parametric efficiency analysis. Therefore, we think that we can have reasonable confidence in the results reported below.

EFFICIENCY DIFFERENCES BETWEEN GROUPS OF RESEARCH PROGRAMS

Our research questions pertaining to inter-university differences can be translated into hypotheses about differences between groups of research programs in terms of central tendency (average or median) of efficiency: comparing the central tendency of different universities or specialization categories can test the impact of the specific university environment or the area of specialization, and efficiency trends over time can be detected by comparing central tendencies of different years.

To address these questions, we subdivide research programs in different subsamples along the following dimensions:

a. *organizing university*: EUR, UvA, VU, UM, KUB, WUR, RUG and KUN;
b. *year*: 1998, 1999, 2000;
c. *specialization type*:[9] A&F = Accounting and Finance, AM = Applied Mathematics, DEV = Development, Growth and Transition, ECO = Econometrics, PUB = Economics of Public Policy, LAB = Applied Labor Economics, MACRO = Macroeconomics, Money and International Issues, M&B = Marketing and Business Economics, S&E = Spatial and Environmental Economics, MICRO = Theoretical and Applied Microeconomics.

[9] This subdivision of specialization types is used in the VSNU assessment.

The standard parametric test for comparing two independently selected samples is the t-test. However, this test builds on normality of the efficiency estimates, which can seem a strong assumption in many cases. For eff1 the Kolmogorov-Smirnov test rejects normality of the efficiency distribution at a significance level of 1%.[10] Hence, the use of the standard t-test seems hardly tenable.

If one wishes to avoid the normality assumption, a number of non-parametric alternatives to the t-test are available. The main advantage of these nonparametric testing tools as compared to their parametric counterparts is that their results are more robust with respect to the underlying distribution of the efficiency values.

The most powerful non-parametric alternative is the Wilcoxon rank sum test for comparing the median of two independent samples, which has about 95% of the power of a t-test under optimal conditions.[11] We use the Wilcoxon statistic to test whether the median efficiency for each of the constructed categories of research programs differs from the median for the other research programs in our sample.

Our Wilcoxon results are reported in Tables 3 and 4. Each table contains for each cell (combination of row category and column category) the average rank ("rank"; lower value reflects higher average efficiency), the number of observations ("number") and the cumulative probability value for the hypothesis that the median efficiency of the programs in that cell equals the median efficiency for the full sample ("p-value"). We reject the null hypothesis of equal median efficiency with respect to the alternative hypothesis that the median efficiency of the subsample is above (below) that of the other programs in our sample if the reported p-value is too low (too high).

The results in Table 3 broadly confirm our earlier findings: KUB and WUR perform systematically better than the other universities; UvA, UM and –to a somewhat lesser extent- KUN are on average less efficient than the rest; research programs were generally most productive in the year 1999, while in 2000 productive efficiency is significantly below the average in the other years. In addition, Table 3 provides interesting information on the performance patterns over years for the different universities. For example, we observe that the efficiency (in terms of average rank) of research programs organized at RUG gradually deteriorates over time, and that, on average and next to KUB and WUR, EUR performs significantly better than the rest in 1999.

Table 4 contains test results for differences over specialization types. We find that the average efficiency of research programs in Econometrics (ECO), Spatial and Environmental Economics (S&E) and Theoretical and Applied Microeconomics (MICRO) is significantly above that of the other research programs. Conversely, research programs in Applied Labor Economics (LAB) and -to a lesser extent- in Accounting and Finance (A&F) and Economics of Public Policy (PUB) seem to perform systematically worse than other programs.

One possible explanation of these rather striking findings is that research programs of different specialization types are not directly comparable, i.e. they face different production possibilities (or, equivalently: relevant performance dimensions have been omitted); however, important efficiency variation within different specialization fields (discussed in greater detail below) somewhat weakens this argument. Another possible interpretation is that (inter)national umbrella organizations function better for one specialization area than for

[10] The same observation holds for eff2 and eff3. One obvious explanation is that the range of possible efficiency values is truncated at unity. The truncated nature of the efficiency measure returns in our further discussion on the impact of program size and external research funding.

[11] Brockett and Golany (1996), for example, advocate this test in the context of nonparametric efficiency analysis.

another. A final explanation is that Dutch universities have a comparative advantage in ECO, S&E and MICRO; the Dutch tradition in e.g. the field of Econometrics provides some support for this position, taking into account possibly beneficial learning effects. Additional research is necessary to investigate these alternative explanations.

Table 4 further provides interesting insights into the university-specific performance in different areas of specialization. For example, we observe substantial efficiency variation in the field of Development, Growth and Transition (DEV; compare UM with WUR), Macroeconomics, Money and International Issues (MACRO; compare UM and KUN with KUB), Marketing and Business Economics (M&B; compare UvA with WUR) and S&E (compare EUR with UvA and VU). These important differences at least suggest that some research programs can learn from the organizational structure of others. Finally, the results reveal that universities may perform rather well in some specialization areas while they do quite badly in others; see e.g. the efficiency variation for EUR, UvA, VU and UM. This implies *inter alia* that a generally poorly performing university should not necessarily operate inefficiently in any specialization domain; see e.g. UvA and UM. This leads us to our main conclusion that considering micro-units of production allows for a more carefully balanced appraisal of the research performance of the different faculties of Economics and Business Management.

Table 3: efficiency over years and universities

		EUR	UvA	VU	UM	KUB	WUR	RUG	KUN	total
1998	rank	116,35	165,39	109,13	142,33	81,89	109,50	105,67	163,50	122,75
	number	20	14	12	9	9	7	6	2	79
	p-value	0,43	1,00	0,30	0,85	0,05	0,35	0,31	0,82	0,72
1999	rank	91,93	119,21	117,67	129,61	65,33	63,07	120,58	138,50	102,73
	number	20	14	12	9	9	7	6	2	79
	p-value	0,03	0,50	0,47	0,68	0,01	0,01	0,52	0,66	0,00
2000	rank	133,53	141,11	118,67	153,06	107,94	101,21	154,17	167,50	131,49
	number	20	14	12	9	9	7	6	2	79
	p-value	0,84	0,89	0,49	0,94	0,31	0,24	0,90	0,84	0,98
total	rank	113,93	141,90	115,15	141,67	85,06	91,26	126,81	156,50	118,99
	number	60	42	36	27	27	21	18	6	237
	p-value	0,25	0,99	0,36	0,97	0,00	0,03	0,69	0,91	

Note: If two or more efficiency values are tied at the same rank, the rank assigned is the average of the ranks which would have been assigned if the scores had differed slightly. This makes that minimal rank is 32,5.

Table 4: efficiency over specialization types and universities

		EUR	UvA	VU	UM	KUB	WUR	RUG	KUN	total
A&F	rank	104,00	158,17	126,83	143,92	111,92		110,67		129,88
	number	6	9	6	6	6	0	3	0	36
	p-value	0,33	0,97	0,66	0,85	0,44		0,44		0,91
AM	rank	129,44	111,67	97,83		109,33				117,86
	number	9	3	3	0	3	0	0	0	18
	p-value	0,73	0,45	0,32		0,43				0,55
DEV	rank			139,58	199,33		51,67	80,83		122,20
	number	0	0	6	3	0	3	3	0	15
	p-value			0,81	0,99		0,05	0,18		0,64
ECO	rank	62,17	111,67	125,67	55,00	49,67				80,83
	number	3	3	3	3	3	0	0	0	15
	p-value	0,08	0,45	0,60	0,06	0,04				0,02
PUB	rank	133,67	169,33							145,56
	number	6	3	0	0	0	0	0	0	9
	p-value	0,74	0,92							0,91
LAB	rank	191,00	160,83	160,00	167,00					167,93
	number	3	6	3	3	0	0	0	0	15
	p-value	0,97	0,95	0,87	0,91					1,00
MACRO	rank	91,42	101,50		164,33	56,17			182,33	114,53
	number	6	3	0	3	3	0	0	3	18
	p-value	0,18	0,35		0,90	0,06			0,96	0,46
M&B	rank	107,61	225,00	149,00	133,83	98,00	73,67	142,33	130,67	123,25
	number	18	3	6	9	9	6	12	3	66
	p-value	0,29	1,00	0,89	0,79	0,20	0,06	0,92	0,65	0,85
S&E	rank	186,33	53,00	42,42			115,33			92,47
	number	3	3	6	0	0	6	0	0	18
	p-value	0,97	0,05	0,00			0,49			0,06
MICRO	rank	73,50	162,33	82,67		32,50	104,58			90,52
	number	6	3	3	0	3	6	0	0	21
	p-value	0,06	0,89	0,19		0,01	0,34			0,03
total	rank	113,93	144,22	115,15	141,67	85,06	91,26	126,81	156,50	118,76
	number	60	36	36	27	27	21	18	6	231
	p-value	0,39	1,00	0,47	0,98	0,01	0,04	0,76	0,93	

Note: Our dealing with tied efficiency values makes that minimal rank is 32,5.

THE IMPACT OF PROGRAM SIZE AND EXTERNAL FUNDING

Information regarding the impact of size and external research funding on the observed research performance can be interesting for program managers, but also for the university and for those in charge of research policy. Specifically, corroboration of the hypothesis that a higher degree of financial support by scientific research funds has a positive impact on research performance (e.g. because it introduces additional, external monitoring on the scientific research process) can stimulate managers of relatively inefficient research programs to attract more external funding.[12] Next, if size has a positive effect on efficiency, this means that relatively inefficient research programs can benefit from 'upsizing', while a negative coefficient can be interpreted as evidence in support of the position that 'small is beautiful'.[13]

Because a categorization of research programs in terms of size and external funding is not readily available, the application of Wilcoxon tests is problematic. Therefore, we employ linear regression techniques in our following exercises. Size is then measured as the total input value in terms of full time equivalent staff members. Next, we measure the degree of external funding as the ratio of total value of WP1 inputs over total input value (i.e. the sum of WP1, WP2 and WP3 inputs).

[12] There is a possible problem of reverse causality in analyzing the impact of external funding as good research performance may also attract external research funding. Therefore, our results concerning the impact of external funding on research efficiency should be considered with sufficient care.

[13] Note that the relationship between program size and program efficiency could be curvilinear – in fact, it seems obvious that there can be an 'optimum' program size. We will take this into account in our regression specifications.

Though the efficiency measure can reasonably be considered continuously distributed, its range is truncated at unity, and several efficiency estimates have that value. We use Tobit regression analysis to handle this problem. Since we estimate the Tobit models through maximum likelihood, we need to specify a priori the conditional distribution of the efficiency measure. In our exercises below we assume the normal distribution. This might seem at odds with our previous discussion on the problematic nature of the normality assumption (see e.g. our results for the Kolmogorov-Smirnov test). Still, that discussion pertained to the *un*conditional distribution of the efficiency measure, and no correction was made for the truncated nature of the observed distribution.

Our results are reported in Tables 5 and 6. In each of these tables, the coefficient (see "coeff.") of "SIZE" gives the first order effect of size on perceived efficiency. Since it can be argued that size affects the performance of research programs nonlinearly, we additionally include the square values of our size measure (see "$SIZE^2$"). Finally, the coefficient of "EXTERNAL" gives the impact of external funding on research performance.

The reported standard errors (see "st. err.") are computed from the second order partial derivatives of the Log likelihood function, and are thus asymptotic standard errors. Asterisks "***", "**" and "*" indicate one-sided significance (see "sign.") at the 1%, 5% and 10% level, respectively.[14]

Table 5 gives the residual effect of size and external funding when controlling for university, year and specialization type, of which the effects are captured through the inclusion of dummy variables.[15] The results in Table 5 fall in line with those in Tables 3 and 4: the universities KUB and WUR are generally most efficient; 1999 has been the most productive year in terms of academic research; Dutch universities achieve the best results in the specialization areas ECO, S&E and MICRO. In addition, we indeed observe a significantly positive effect of external funding on research efficiency.

Still, we do not find significant evidence in support of a size effect. One interpretation is that there simply is no such effect. However, other explanations are equally possible. For example, the inclusion of too many inter-correlated independent variables can reduce the significance of the observed effect. Or, perhaps even more likely in view of our earlier findings regarding efficiency variation over specialization types, the specific size effect could depend on the particular area of specialization.

The results in Table 6 allow us to check these alternative interpretations. Since we can reasonably ignore the year effects when investigating the influence of size and external funding, we merely control for the possible impact of the organizing university. In particular, we include dummy variables denoting whether the university efficiency was identified generally above central tendency ("UNIVERSITY > AVG"; capturing KUB and WUR) or below central tendency ("UNIVERSITY < AVG"; capturing UvA, UM and KUN) in our Wilcoxon rank sum analysis.

The table contains results for 11 regressions. Our first regression exercise computes results for the pooled set of research programs (see the column "TOTAL"). The significant

[14] The test results build on asymptotic properties of the maximum likelihood estimators. This calls for using larger data sets to potentially obtain even better results; see also our suggestion in the concluding section to extend the current analysis to a cross-country analysis.

[15] To avoid the so-called 'dummy variable trap', we exclude dummies for the categories UvA (university type), 2000 (year) and LAB (specialization type). Hence, these omitted categories act as reference groups and all effects are relative to these groups. This also makes that the results in Table 5 are not fully comparable to our Wilcoxon results, where the benchmark for each category of research programs was the group of remaining programs.

coefficients of the university and external funding parameters have the expected sign. Once more, however, we do not find statistical evidence of a size effect when considering all specialization types together.

The other regressions pertain to specific specialization categories (see the columns “A&F”; “AM”; “DEV”; “ECO”; “PUB”; “LAB”; “MACRO”; “M&B”; “S&E”; “MICRO”). Unsurprisingly, significant coefficients for “UNIVERSITY > AVG” and “UNIVERSITY < AVG” are positive and negative, respectively.

As for external funding, the hypothesized positive effect is significantly confirmed for not less than 6 out of 10 specialization areas: DEV, ECO, MACRO, M&B, S&E and MICRO. In these fields, relatively inefficient research programs that are characterized by a low degree of external funding might substantially improve their productive efficiency by attracting additional external funding.

Finally, the results in Table 6 do reveal size effects. Specifically, for the categories AM and MACRO a positive first-order effect (see the coefficients of “SIZE”) combined with a negative second order effect (see the coefficients of “SIZE²”) suggests that productive efficiency is significantly increasing in size, but to an ever decreasing extent; the size effect becomes negative for very large research programs.[16] A similar, though less outspoken, effect is observed for DEV and S&E. The opposite relationship seems to hold for research programs in the area A&F; here, the suggested relationship between efficiency and size has a ‘U’-shape, which means that the ‘optimal’ size is either very small or very large.

Table 5: Tobit regression results; all variables included

		TOTAL (237 observations)		
category		**coeff.**	**st.err.**	**sign.**
	intercept	0,1680	0,1620	
	SIZE	0,0022	0,0090	
	SIZE²	0,0000	0,0002	
	EXTERNAL	0,4652	0,2073	**
Year	1998	0,0712	0,0613	
	1999	0,1709	0,0611	***
Specialization	A&F	0,1556	0,1131	*
	AM	0,1169	0,1312	
	DEV	0,1779	0,1421	
	ECO	0,4272	0,1379	***
	PUB	0,0630	0,1556	
	MACRO	0,2535	0,1360	**
	M&B	0,1722	0,1078	*
	S&E	0,2848	0,1363	**
	MICRO	0,3244	0,1310	***
University	EUR	0,2072	0,0953	**
	VU	0,1330	0,0964	*
	UM	0,0414	0,1130	
	KUB	0,3712	0,1171	***
	WUR	0,2690	0,1275	**
	RUG	0,1393	0,1326	
	KUN	-0,1189	0,1740	
	log likelihood	-131,471		

[16] The coefficients of “SIZE” and “SIZE²” actually allow for estimating the ‘optimal’ program size for the different specialization domains. Still, these estimates would be rather imprecise.

Table 6: Tobit regression results; different specialization types

		TOTAL	A&F	AM	DEV	ECO	PUB	LAB	MACRO	M&B	S&E	MICRO
# observations		237	36	18	15	15	9	15	18	66	18	21
intercept	coeff.	0,6049	1,0933	-1,1501	-0,0598	0,2064	-0,1733	0,3454	-0,7390	0,4585	-1,8879	1,2134
	st. err.	0,0960	0,2315	0,8916	0,7418	0,7183	0,6142	0,1438	0,4843	0,1806	1,6194	0,4579
	sign.	***	**					**	*	***		***
UNIVERSITY > AVG	coeff.	0,1906	-0,0793	0,0132	0,9126	0,5399			0,5074	0,2650	0,8534	0,1352
	st. err.	0,0731	0,1676	0,2301	0,4480	0,3505			0,1614	0,1303	0,7592	0,2280
	sign.	***			**	*			***	**		
UNIVERSITY < AVG	coeff.	-0,1646	-0,2458	0,2955	-0,4327	0,3527	-4,4349	0,0670	0,1409	-0,1213	1,3327	-1,3288
	st. err.	0,0624	0,1293	0,9232	0,2192	0,2772	3,4650	0,1069	0,1431	0,1230	1,6108	0,4386
	sign.	***	**		**							***
SIZE	coeff.	0,0052	-0,0376	0,1499	0,1011	0,0115	0,1128	-0,0153	0,1245	0,0150	0,2515	-0,0676
	st. err.	0,0079	0,0195	0,0792	0,0889	0,0504	0,0886	0,0279	0,0421	0,0146	0,1671	0,0568
	sign.		**	**					***		*	
SIZE²	coeff.	-0,0001	0,0006	-0,0029	-0,0037	0,0000	-0,0005	0,0004	-0,0032	-0,0003	-0,0060	0,0014
	st. err.	0,0001	0,0003	0,0018	0,0026	0,0010	0,0011	0,0006	0,0009	0,0003	0,0051	0,0011
	sign.		**	*	*				***			
EXTERNAL	coeff.	0,4086	0,3650	0,6623	1,8982	1,9434	3,2027	0,1208	3,0301	0,8728	1,0551	4,1067
	st. err.	0,1911	0,9420	0,6529	0,6211	1,0705	3,3274	0,4891	0,8287	0,6569	0,6913	1,6056
	sign.	**			***	*			***	*	*	**
log likelihood		-143,1330	-16,3265	-6,4235	1,2992	-0,7800	-4,9544	4,3765	-0,1587	-37,8059	-11,0931	-6,5234

Note: WUR and KUB (captured in 'UNIVERSITY>AVG') do not organize research programs in the areas PUB and LAB.

5. CONCLUDING DISCUSSION

We have analyzed the productive efficiency of research in Economics and Business Management at Dutch universities. Our approach is original in at least two respects: we concentrate on micro-units of research production (*in casu* research programs) rather than on conventionally employed macro-units (such as faculties or departments); and we employ a non-parametric efficiency assessment methodology that starts from a characterization of academic research objectives rather than the (typically unknown) technological possibilities.

The main conclusions of our empirical study can be summarized as follows:

a. considerable differences in research performance over universities make it hard to deny the impact of the university environment on academic research performance; in particular the efficiency of Tilburg University and Wageningen University is significantly higher than that of other universities;
b. we observe important efficiency variation over time; efficiency is higher in 1999 than in 1998, which is followed –rather surprisingly- by a dramatic efficiency deterioration in 2000;
c. efficiency values vary substantially over specialization areas; Dutch universities appear to have a comparative advantage in Econometrics, Spatial and Environmental Economics and Theoretical and Applied Microeconomics;
d. we find statistical evidence in support of size effects in several specialization areas; the direction of the relationship between size and productive efficiency seems to depend on the specific category of specialization;
e. we obtain strong and persistent corroboration of a positive relationship between efficiency of academic research and the degree of financial support by scientific research funds.

Our results illustrate the usefulness of focusing on micro-units of research; see e.g. the efficiency variation over specialization domains (between and within universities) and the specialization-dependent size effects. Generally, such practice allows for a more carefully balanced evaluation of research performance. In addition, and perhaps even more interestingly, these detailed results can provide helpful input to research managers in designing their research policy. For example, research programs/universities that perform poorly can learn from the organizational structure of comparable research programs/universities that perform generally well. Further, our results concerning the impact of size and external funding on research performance may reveal direct policy directions for the managers of inefficient research programs.

To conclude, we point out a number of interesting avenues for further research.

a. *Other and/or larger samples*. Evidently, it would be interesting to apply our methodology for evaluating research institutes in other countries, and to contrast these results with those presented in the current study. Alternatively, data for research institutes in different countries can be pooled to perform a cross-country analysis, which can expose significant differences over countries and which may even suggest guidelines for national policy makers. Furthermore, larger data sets can deepen our insights regarding the determinants of academic research performance.
b. *Additional determinants of research efficiency*. Limited availability of objective data has forced us to restrict our attention to a rather limited set of possible determinants of research performance (university environment, area of specialization, external funding and program size). It would be interesting to further investigate why some research programs/universities perform systematically better (or worse) than comparable research programs/universities, e.g. by contrasting organizational features such as incentive schemes or the degree of international networking. Or, we could study why research programs in some areas of specialization steadily outperform research programs in other specialization fields, e.g. by relating efficiency values to the functioning of (inter)national umbrella organizations for the different specialization fields. Such research would generate an even better insight into the specificities of research efficiency, which would eventually benefit research performance in general.
c. *Data issues*. Recall that our methodology is sensitive to errors-in-the-data and omitted performance dimensions. Although we have somewhat mitigated these problems by using analytical tools that account for possible outlier behavior, and by computing efficiency results for different sets of input and output dimensions, this remains an important issue for further research. At least, it calls for a well-conceived data collection procedure, taking special care of the data quality and the selection of (standardized) inputs and outputs. The data set used in the current study is indeed well standardized and has been subjected to some scrutiny for correctness and consistency. For low-quality data sets, it is recommendable to develop methodological extensions that satisfactorily deal with this kind of data problems; see e.g. Grosskopf (1996) for a survey of tools that are currently available in the non-parametric literature.
d. *Other application settings*. Our approach starts from micro-units of research production and a (minimal) specification of research objectives to end up with summarizing measures of productive efficiency. It should be easy to adapt this method to other academic research disciplines. In addition, we believe that our approach is readily applicable to research activities in the business sector; see e.g. Hauser (1998) for general guidelines pertaining to the selection of appropriate research performance metrics for successful R&D management. Finally, our approach can be useful for quantifying the success of universities in transferring their technology to the business community, a subject that has become topical in recent years; see e.g. Mowery and Shane (2002).

REFERENCES

S. AFRIAT (1972): Efficiency Estimation of Production Functions, *International Economic Review* 13, 568-598.

R. BANKER (1993): Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation, *Management Science* 39, 1265-1273.

P. BOGETOFT (1996): DEA on Relaxed Convexity Assumptions, *Management Science* 42, 457-465.

P.L. BROCKETT AND B. GOLANY (1996): Using Rank Statistics for Determining Programmatic Differences in DEA, *Management Science* 42, 466-472.

R.G. COOPER (1996): *Winning at New Products: Accelerating the Process from Idea to Launch*, Addison-Wesley, Reading.

W.W. COOPER, L. M. SEIFORD, AND K. TONE (2000): *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software,* Kluwer Academic Publishers, Dordrecht.

R. CORDERO (1990): The Measurement of Innovation Performance in the Firm: An Overview, *Research Policy* 19, 185-192.

G. DEBREU (1951): The Coefficient of Resource Utilization, *Econometrica* 19, 273-292.

R. FÄRE, S. GROSSKOPF AND C.A.K. LOVELL (1994): *Production Frontiers*, Cambridge University Press, Cambridge.

M.J. FARRELL (1957): The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society Series A* 120, 253-281.

A. GRIFFIN, AND J. R. HAUSER (1996): Integrating Mechanisms for Marketing and R&D, *Journal of Product Innovation Management* 13, 191-215.

S. GROSSKOPF (1996): Statistical Inference and Non-Parametric Efficiency: A Selective Survey, *Journal of Productivity Analysis* 7, 161-176.

G. HANOCH AND M. ROTHSCHILD (1972): Testing Assumptions of Production Theory: A Nonparametric Approach, *Journal of Political Economy* 80, 256-275.

J.R. HAUSER (1998): Research, Development and Engineering Metrics, *Management Science* 44, 1670-1689.

M. G. KOCHER, M. LUPTÁCIK AND M. SUTTER (2001): *Measuring Productivity of Research in Economics. A Cross-Country Study Using DEA*, Department of Economics Working Paper Series 77, Vienna University of Economics.

P. KORHONEN, R. TAINIO AND J. WALLENIUS (2002): Value Efficiency Analysis of Academic Research, *European Journal of Operational Research*, forthcoming.

D.C. MOWERY AND S. SHANE (2002): Editors of 'Special Issue on University Entrepreneurship and Technology Transfer', *Management Science* 48.

L. SIMAR AND P. WILSON (2000): Statistical Inference in Nonparametric Frontier Models: The State of the Art, *Journal of Productivity Analysis* 13, 49-78.

J. G. THURSBY (2000): What Do We Say about Ourselves and What Does It Mean? Yet Another Look at Economics Departments Research, *Journal of Economic Literature* 38, 383-404.

H.R. VARIAN (1984): The Non-Parametric Approach to Production Analysis, *Econometrica* 52, 279-297.

H.R. VARIAN (1990): Goodness-of-Fit in Optimizing Models, *Journal of Econometrics* 46, 125-140.

APPENDIX A: ALPHABETICAL LIST OF INTERNATIONAL 'TOP' JOURNALS

Academy of Management Journal; Academy of Management Review; Administrative Science Quarterly; American Economic Review; American Journal of Agricultural Economics; Brooking Papers on Economic Activity; California Management Review; Demography; Economic Geography; Econometrica; Economic Journal; Environment and Planning A; European Economic Review; European Journal of Operational Research; Foreign Affairs; Harvard Business Review; Human Resource Management; Industrial and Labor Relations Review; Industrial Relations; International Economic Review; International Journal of Urban and Regional Research; International Organization; Journal of Business; Journal of Business and Economic Statistics; Journal of Conflict Resolution; Journal of Consumer Research; Journal of Development Economics; Journal of Econometrics; Journal of Economic Perspectives; Journal of Economic Theory; Journal of Economic Literature; Journal of Environmental Economics and Management; Journal of Finance; Journal of Financial Economics; Journal of Health Economics; Journal of Human Resources; Journal of International Business Studies; Journal of International Economics; Journal of Labor Economics; Journal of Law and Economics; Journal of Law, Economics and Organization; Journal of Management; Journal of Marketing; Journal of Marketing Research; Journal of Monetary Economics; Journal of Money, Credit and Banking; Journal of Public Economics; Journal of Political Economy; Journal of Product Innovation Management; Journal of Risk and Uncertainty, Journal of Urban Economics; Management Science; Mathematics of Operations Research; MIS Quarterly; Operations Research; Organization Science; Organizational Behavior and Human Resource Decision Processes; Oxford Bulletin of Economics and Statistics; Population and Development Review; Population Studies; Quarterly Journal of Economics; RAND Journal of Economics; Regional Studies; Review of Economic Studies; Review of Economics and Statistics; Review of Financial Studies; Sloan Management Review; Strategic Management Journal; Urban Studies; World Development.

APPENDIX B: AVERAGE EFFICIENCY OF INDIVIDUAL RESEARCH PROGRAMS (1998-2000)

program	eff1	eff2	eff3	program	eff1	eff2	eff3
EUR 01	100,00%	100,00%	100,00%	VU 07	93,42%	100,00%	100,00%
EUR 02	61,58%	67,30%	79,67%	VU 08	44,15%	45,07%	44,15%
EUR 03	44,70%	97,78%	93,89%	VU 09	75,01%	75,01%	75,01%
EUR 04	27,02%	99,68%	31,12%	VU 10	70,90%	75,92%	85,32%
EUR 05	31,10%	62,13%	40,21%	VU 11	27,36%	31,96%	45,27%
EUR 06	100,00%	100,00%	100,00%	VU 12	75,16%	78,25%	100,00%
EUR 07	59,74%	76,47%	59,74%	UM 01	35,49%	81,67%	48,80%
EUR 08	53,32%	62,98%	61,65%	UM 02	39,12%	46,49%	68,19%
EUR 09	96,25%	100,00%	100,00%	UM 03	61,64%	67,03%	61,64%
EUR 10	48,81%	49,52%	48,81%	UM 04	38,02%	48,08%	56,78%
EUR 11	93,55%	100,00%	93,55%	UM 05	91,37%	100,00%	91,37%
EUR 12	68,37%	87,82%	68,37%	UM 06	25,47%	60,32%	32,82%
EUR 13	84,40%	89,80%	84,40%	UM 07	76,53%	93,13%	88,98%
EUR 14	18,51%	24,73%	18,51%	UM 08	53,10%	56,29%	60,00%
EUR 15	68,04%	70,53%	68,04%	UM 09	42,70%	52,89%	63,41%
EUR 16	83,36%	100,00%	89,35%	KUB 01	70,10%	75,84%	70,10%
EUR 17	48,18%	59,84%	63,68%	KUB 02	95,77%	100,00%	95,77%
EUR 18	63,25%	91,38%	66,37%	KUB 03	60,08%	95,92%	60,08%
EUR 19	47,61%	65,66%	78,16%	KUB 04	98,38%	100,00%	99,78%
EUR 20	85,14%	85,60%	94,87%	KUB 05	84,91%	87,63%	85,00%
UvA 01	66,71%	97,57%	66,71%	KUB 06	100,00%	100,00%	100,00%
UvA 02	69,28%	91,31%	79,77%	KUB 07	70,29%	70,29%	93,19%
UvA 03	41,37%	67,47%	57,43%	KUB 08	70,65%	100,00%	90,65%
UvA 04	47,14%	65,93%	47,14%	KUB 09	58,97%	77,55%	65,16%
UvA 05	22,81%	49,84%	30,65%	WUR 01	81,06%	85,45%	82,63%
UvA 06	13,77%	16,08%	13,77%	WUR 02	94,30%	100,00%	100,00%
UvA 07	64,02%	69,47%	76,11%	WUR 03	71,02%	74,69%	76,47%
UvA 08	92,77%	100,00%	100,00%	WUR 04	70,66%	100,00%	87,79%
UvA 09	67,01%	71,48%	81,43%	WUR 05	46,51%	50,04%	97,22%
UvA 10	37,71%	59,63%	81,66%	WUR 06	75,33%	89,36%	80,42%
UvA 11	45,26%	78,16%	47,07%	WUR 07	87,37%	100,00%	87,37%
UvA 12	77,02%	78,53%	100,00%	RUG 01	81,92%	89,56%	82,96%
UvA 13	37,19%	80,83%	47,65%	RUG 02	65,43%	100,00%	65,43%
UvA 14	40,55%	94,44%	40,55%	RUG 03	69,95%	84,38%	81,98%
VU 01	77,74%	77,74%	77,74%	RUG 04	29,14%	48,42%	29,14%
VU 02	61,30%	75,09%	62,38%	RUG 05	34,40%	54,89%	41,12%
VU 03	45,88%	52,35%	62,15%	RUG 06	69,39%	79,34%	73,84%
VU 04	31,42%	47,02%	85,14%	KUN 01	31,38%	37,71%	49,66%
VU 05	68,77%	68,77%	83,87%	KUN 02	58,38%	87,27%	73,78%
VU 06	100,00%	100,00%	100,00%				